Argonne NATIONAL LABORATORY

INCITE Proposal Writing Tips

Ray Bair, Chief Computational Scientist, CELS
Argonne National Laboratory and The University of Chicago

June 20, 2008

INCITE Proposal Writing Tips

- INCITE at ALCF
- What is INCITE good for? …Not so good for?
- INCITE Review Process
- Science Story Tips
- User's View of ALCF Systems
- User Agreements
- PI Obligations
- Additional Q&A

U.S. Department of Energy
Office of Science

Crunching Since 2004

INCITE
Innovative and Novel Computational Impact on Theory and Experiment

- Solicits large computationally intensive research projects
  - to enable high-impact scientific advances
- Open to all scientific & engineering researchers and organizations
  - including universities, laboratories and companies
- Provides large computer time & data storage allocations
  - to a small number of projects for 1-3 years

Does not provide funds for research

2008 INCITE Projects on Blue Gene/P

111 Million Hours on Blue Gene/P at ALCF

Argonne NATIONAL LABORATORY

New Projects	
Project	**Hours**
Computational Protein Structure Prediction and Protein Design	12,000,000
Computational Nuclear Structure	10,000,000
Large-scale Simulations of Cardiac Electrical Activity	846,720
Modeling the Rheological Properties of Concrete	750,000
Study of Buoyancy-Driven Turbulent Nuclear Burning and Validation of Type Ia Supernova Models	21,000,000
Molecular Simulation of Complex Chemical Systems	750,000
Blue Gene/P Plan 9 Measurements on Large Scale Systems	1,000,000
Massively Parallel Simulation of Combustion in Gas Turbines	4,000,000
Predictions of Thermal Striping in Sodium Cooled Reactors	5,000,000
Lattice QCD	19,600,000
High Resolution Global Simulation of Plasma Microturbulence	2,000,000
Climate-Science Computational End Station Development and Grand Challenge Team	1,000,000
Kinetics and Thermodynamics of Metal and Complex Hydride Nanoparticles	1,000,000
Simulation and Modeling of Synuclein-based 'Protofibril Structures' as a Means of Understanding the Molecular Basis of Parkinson's Disease	1,200,000
Performance Evaluation and Analysis Consortium End Station	4,000,000

Renewal Projects	
Project	**Hours**
Molecular Simulations of Surfactant Assisted Aqueous Foam Formations	4,000,000
High Fidelity LES Simulations of an Aircraft Engine Combustor to Improve Emissions and Operability	1,377,000
Reactor Core Hydrodynamics	14,000,000
Water in Confined Spaces	6,000,000
Gating Mechanism of Membrane Proteins	1,500,000

A LITTLE SKEPTICISM CAN BE HEALTHY

- These are my opinions based on experience with INCITE and other HPC proposal programs. Your mileage may vary.
- Help and advice is available from many sources, e.g., ALCF staff, DOE Program Managers, colleagues with INCITE projects.
- The story behind each science campaign is different. Use your best judgment when writing your proposal.

What is INCITE good for?

- High impact science and engineering with specific objectives
- Computationally intensive runs that you cannot do anywhere else
- Jobs that can use at least 25% of the system for hours to weeks
- Campaigns requiring millions of CPU hours
- Computations that are efficient on ALCF's system

What is INCITE not so good for?

- Small projects, e.g., < 5,000,000 CPU hours
- Jobs that never use a large number of processors
- Computations that cannot be check-pointed
- Long interactive computations
- Work that must be co-scheduled with distributed resources
- Datasets bigger than the file system

Some of the above characteristics are negotiable, so make sure to discuss atypical requirements with ALCF

INCITE Proposal Review

- Scientific Discipline Peer Review
  - Scientific / engineering quality
  - Proposed impact of the project
  - Ability of the PI and team
  - Computational plan
  - Relation to the Office of Science mission-related research
- Computational Readiness Review
  - Reasonableness and appropriateness of resource request
  - Appropriateness of approach
  - Technical readiness - has code run at scale on target system?
  - Progress in previous year (for renewals)

IT IS A SMALL WORLD…

- DOE program managers interact on many occasions, within the Office of Science, within DOE, and with other science agencies
- Advice is often sought and given
- So it could help to let the science program that funds your work know how significant your planned INCITE allocation will be to your work and their program

Have your elevator speech ready

Science & Engineering Story Tips

Argonne NATIONAL LABORATORY

- Audience
  - Computational science savvy senior scientists / engineers, faculty, and program managers
  - Not everyone will be well versed in your approach
- Story Elements
  - What the problem is, and its significance
  - Key objectives, key simulations/computations, project milestones
  - Approach to solving the problem, its challenging aspects, preliminary results
  - Impact of a successful computational campaign - the big picture
  - Reasons why it is important to carry out this work now

INCITE has become quite competitive

Administrative Tips

- Project Title
  - Pick a title you will be proud of seeing in many, many places
  - Be succinct - no kidding!

- Project Summary
  - (4a) 2-paragraphs that cover the main points of the story
  - (4b) 2-sentences suitable for the public (e.g., *Science News*, *Scientific American*, web, PowerPoint)

Note: the PI is unlikely to also be the Institutional Contact

Computational Resource Request

- How many years will your project need (1-3)?
- Things that will slow you down the first year
  - Porting and code development
  - Learning to use a new center
  - User agreements for all the institutions and people involved
  - Paperwork for proprietary use
- Mind the units!
  - Processor (Core) Hours - for the system you will be running on
    - On Blue Gene you are charged for all cores in your partition
    - Large partitions are in increments of a rack (1,024 nodes, 4,096 cores)
  - Disk Storage - in gigabytes for both Home and Scratch space
  - Mass (Tape) Storage - in gigabytes or terabytes (specify)

Alternate vs. Multiple Facilities

- If your project needs a single primary resource
  - Identify the primary resource in (5a)
  - Identify an alternate resource in (5b) – optional
- If your project needs multiple primary resources
  - Identify each of the primary resources in (5a)
    - By adding resources for the same year
  - Identify any alternate resources in (5b)
  - Describe your need for multiple resources in sections (8) and readiness in (9)

Computational Approach

- Essential to show experience and credibility in this section
- Programming Languages, Libraries and Tools used
  - Check that what you need is available on the system
- Description of underlying formulation
  - Can be short or long, depending on how well known your application is
  - Don't assume reviewers know all the codes

Computational Plan

- Application Packages
  - List of all the software application packages/suites to be used

    Note: Long lists may reduce credibility
  - What will be used to set up computations (and where)?
  - What are the codes for the main simulation/modeling?
  - What will be used to analyze results (and where)?

Application Credentials

- Port your code before submitting the proposal
  - Check to see if someone else has already ported it
  - Request a start up account if needed
- It is very hard to embarrass a 163,480 processor system
  - Prove application scalability in your proposal
  - If at all possible, run example cases at full scale
  - If you cannot show proof of runs at full scale, then provide a very tight story about how you will succeed at full scale

Available BG/P Applications

Application	Field	Perf Info	Globally Available	Makefile or Help
FLASH	Astrophysics	Y	N	Y
MILC,CPS	LQCD	Y	Y	Y
Nek5000	CFD, nuclear energy	Y	N	Y
Rosetta	Protein Structure		N	N
ANGFMC	Nuclear Structure		N	N
Qbox	Phys Chem	Y	N	N
LAMMPS	Molecular Dynamics		Y	Y
CHARMM	Molecular Dynamics	Y	Y	Y
NAMD	Molecular Dynamics	Y	Y	Y
AVBP	Combustion	Y	N	N
GTC	Fusion	Y	N	Y
Allstar	Life Science		N	N
CPDM, CP2K	Molecular Dynamics		N	N
CCSM3	Climate		N	Y
HOMME	Climate	Y	N	Y
WRF	Climate		N	Y
Amber9	Molecular Dynamics		Y	Y
NWChem	Chemistry		N	Y
enzo	Astrophysics		N	Y
Falkon	Computer Science/HTC		Y	Y

BG/P Programming Models

- Parallel Programming System
  - MPI (MPICH2) is the work horse on Blue Gene
    ARMCI/Global Arrays is available
  - OpenMP on nodes; 1 thread per core
  - Some groups rolled their own at lower level, e.g., QCD
- Special needs? Inquire
  - e.g., Python, custom kernel

Tools/Libraries/Packages Supported by ALCF

Argonne NATIONAL LABORATORY

PETSc	HPC Toolkit	mercurial
BLAS	fftw2, fftw3	svn
LAPACK	LIBGOTO	gcc
GOTOBLAS	SCALAPACK	
mass/massv	p3dfft	
ESSL	hypre	
HDF5	MUMPS	
PNETCDF	spooles	
Totalview	SuperLU	
TAU	Coreprocessor	

Computational Campaign

- Describe the kind of runs you plan with your allocation
  - L exploratory runs using M nodes for N hours
  - X big runs using Y nodes for Z hours
  - P analysis runs using Q nodes for R hours
- Big runs often have big output
  - Show you can deal with it and understand the bottlenecks
  - Understand the size of results, where you will analyze them, and how you will get the data there

Parallel Performance

Direct evidence is required

Argonne NATIONAL LABORATORY

- Pick the approach relevant to your work and show results
  - Strong Scaling Data
    - Increase resources (nodes) while doing the same computation
  - Weak Scaling Data
    - Increase problem size as resources are increased
- Performance data should support the required scale
  - Use similar problems to what you will be running
  - Show that you can get to the range of processors required
  - Best to run on the same machine, but similar size runs on other machines can be useful
  - Describe how you will address any scaling deficiencies
- Be aware of scaling data from other groups and literature

I/O Requirements

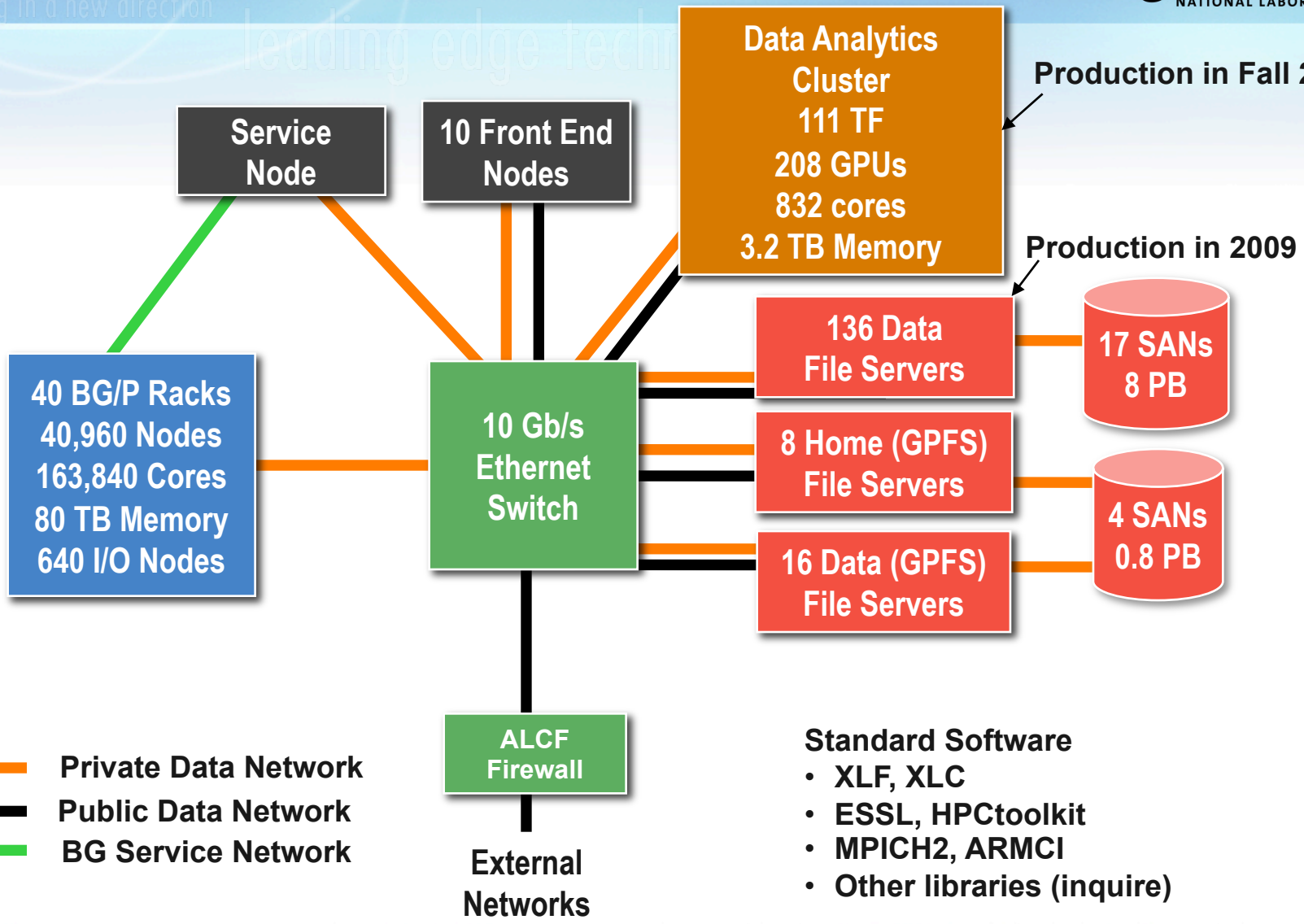
- Restart I/O - Application initiated program restart data
  - I/O technique used, e.g., MPI I/O, HDF5, raw
  - Number of processors doing I/O, number of files
  - Sizes of files and overall dump
  - Periodicity of the checkpoint process
- Analysis I/O - Application written files for later analysis
  - I/O technique used, e.g., pNetCDF, pHDF5
  - Number of processors doing I/O, number of files
  - Sizes of files and overall dump
- Archival I/O - Data archived for later use/reference
  - Number and sizes of files
  - Retention length
  - If archived remotely, the transport tool used, e.g., GridFTP

Blue Gene/P "Intrepid" System

2008: 8 Racks Production, 32 Racks Early Science

2009: 40 Racks Production

Argonne NATIONAL LABORATORY

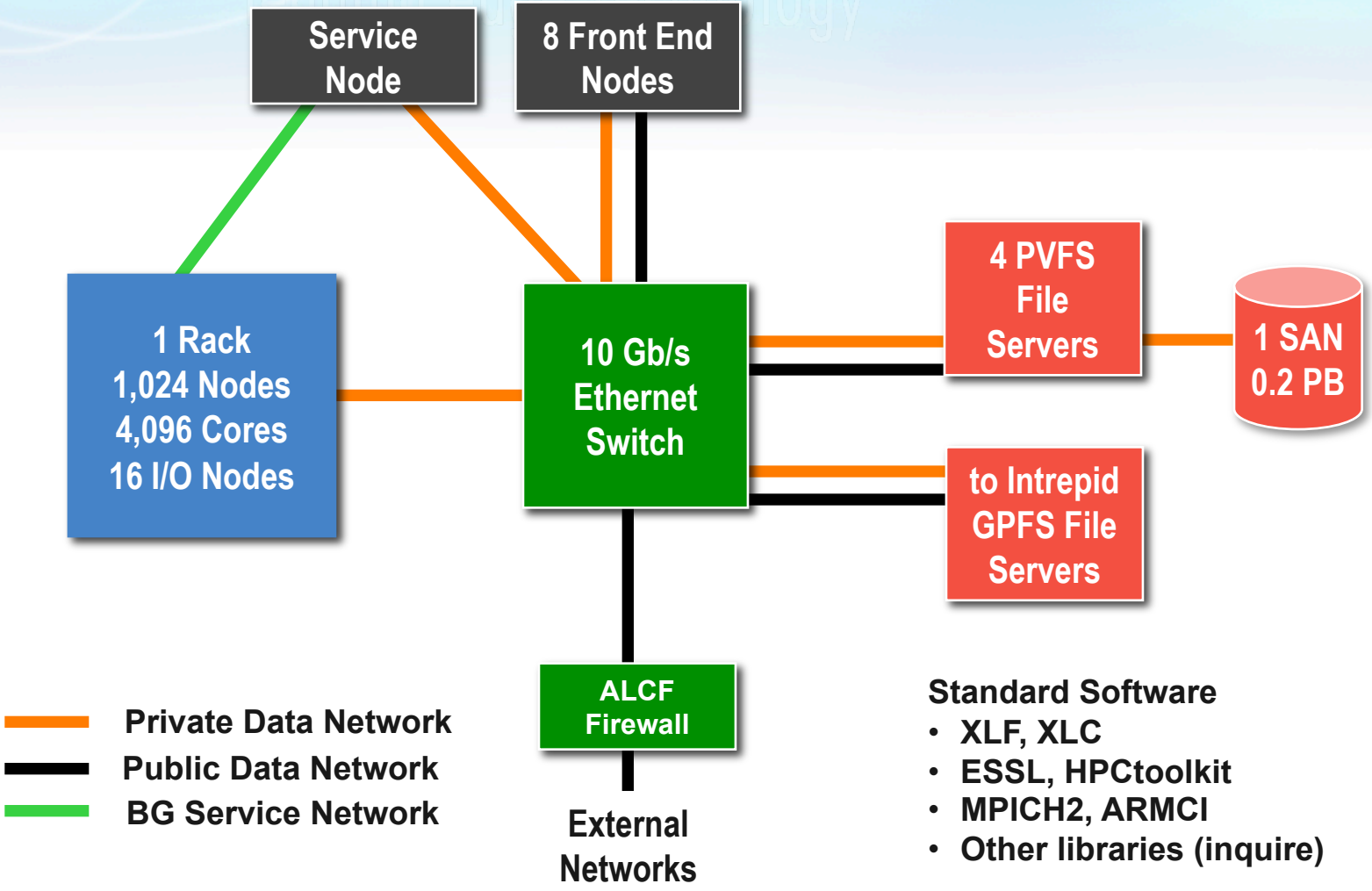


Standard Software

- XLF, XLC
- ESSL, HPCtoolkit
- MPICH2, ARMCI
- Other libraries (inquire)

Surveyor BG/P System

Porting, Testing, Development



Standard Software

- XLF, XLC
- ESSL, HPCtoolkit
- MPICH2, ARMCI
- Other libraries (inquire)

User Agreements

- Nonproprietary Research
  - Substantial results published in open literature or reports
  - Federally-funded user agreement – if any US Federal funds
  - Privately-funded user agreement – for industry, etc.
  - INCITE Program User Agreement is not negotiable
- Proprietary Research is permitted
  - Results retained by researcher or their organization
  - Full cost recovery. Data protection considerations
  - Carefully read DOE guidelines, and begin discussions early
- Process
  - Each institution must sign INCITE Program User Agreement - by one who can commit the institution, e.g., attorneys
  - Every user must have their own account, and must individually complete a facility user agreement

KNOW WHAT KIND OF INFORMATION YOUR PROJECT USES AND GENERATES

Argonne NATIONAL LABORATORY

- Laws regulate what can be done on these systems
  - Also LCF systems have cybersecurity plans that bound the types of data that can be used and stored on them
  - (see www.alcf.anl.gov/support/usingALCF/docs/dataprivacy.php)
- Some kinds of information we cannot have at ALCF
  - Personally Identifiable Information (PII)
  - Information requiring an export control license
  - Classified Information or National Security Information
  - Unclassified Controlled Nuclear Information (UCNI)
  - Naval Nuclear Propulsion Information (NNPI)
  - Information about development of nuclear, biological or chemical weapons, or weapons of mass destruction
- Inquire if you are unsure or have questions

PI Obligations

- Provide quarterly status updates (on supplied template)
  - Milestone reports
  - Publications, awards, journal covers, presentations, etc. related to the work
- Provide highlights on significant science/engineering accomplishments as they occur
- Submit annual renewal request
- Complete annual surveys
- Encourage your team to be good citizens on the computers
- Use the resources for the proposed work
- Let ALCF know about problems and issues

When you win the INCITE award

- Notice comes from DOE INCITE Program Manager, Barbara Helland
- Welcome and startup information comes from Argonne
  - Agreements to sign. How to get accounts.
  - Startup workshop is scheduled ASAP
- User Support is geared to help you succeed
- Catalysts provide expert-to-expert assistance
- Performance Engineering and Data Analytics team help you get the most from your allocation

INCITE 2009

Argonne NATIONAL LABORATORY

Call for proposals	May 16
Proposals due	August 11
Instructions	http://hpc.science.doe.gov

ALCF Information	www.alcf.anl.gov
Getting Started on BG/P	www.alcf.anl.gov/support/gettingstarted/
ALCF Service desk	support@alcf.anl.gov 1-866-508-9181

Additional Questions?